Theory and Practice of Data Assimilation in Ocean Modeling

Robert N. Miller
College of Oceanic and Atmospheric Sciences
Oregon State University
Oceanography Admin. Bldg. 104
Corvallis, OR 97331-5503
Phone: (541) 737-4555   Fax: (541) 737-2064   Email: miller@coas.oregonstate.edu

LONG-TERM GOALS

The long-range goal of this project is to combine computational models with observational data to form the best picture of the ocean as an evolving system, and use that picture to understand the physical processes that govern the ocean's behavior. Oceanic observations are sparse and models are limited in accuracy, but judiciously constructed combinations of data and model output have the potential to form a quantitative description of the state of the ocean that is superior to any based on either models or data alone. Along with the goals of analysis and prediction, we seek reliable estimates of the errors in our results. We expect our results to have implications beyond data assimilation. In particular, we believe this research will lead to enhanced understanding of the implications of nonlinearity and randomness for predictability of the ocean and atmosphere.

In keeping with our goal of providing reliable error estimates for our data assimilation products, we seek to develop efficient methods for estimating useful statistical measures of errors in stochastic forecast models. Since the probability density functions (PDFs) of nonlinear stochastic models are not, in general, Gaussian, we must find methods for forecast evaluation based on information about the particular PDF generated by the model.

Since our goal is the development of practical analysis and forecast systems for the ocean, we want to solve remaining scientific problems involved in transition from data assimilation experiments tuned to specific models and data sets to operational analysis and prediction on a research basis. This will involve rigorous quantification of the information content of each data set, as well as quality control, a problem with which the ocean modeling community has limited experience.

OBJECTIVES

The principal objective of this project is the development, implementation and evaluation of practical data assimilation methods for regional to basin scale ocean models. Since the data assimilation methods that give the most and best information are highly resource intensive, and often not practical for use with detailed models, we are particularly interested in the price paid in terms of accuracy and confidence for using economical but suboptimal data assimilation methods.

Explicit estimation of PDFs in the high-dimensional state spaces that characterize ocean models is not practical, and probably would not be useful if it could be achieved, but useful approximations can be calculated from Monte-Carlo experiments, by virtue of the fact that the number of truly independent

		Form Approved OMB No. 0704-0188

Report Documentation Page

1. REPORT DATE **30 SEP 2007**	2. REPORT TYPE **Annual**	3. DATES COVERED **00-00-2007 to 00-00-2007**
4. TITLE AND SUBTITLE **Theory And Practice Of Data Assimilation In Ocean Modeling**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Oregon State University,College of Oceanic and Atmospheric Sciences,Oceanography Admin. Bldg. 104,Corvallis,OR,97331**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**		
13. SUPPLEMENTARY NOTES **code 1 only**		
14. ABSTRACT		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT **Same as Report (SAR)**	18. NUMBER OF PAGES **7**	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**			

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

degrees of freedom in practical models is much smaller than the dimension of the state vector. This is the motivation for the ensemble methods that are used by major weather centers to evaluate forecast reliability. Dynamical systems theory gives us methods for defining candidate low dimensional structures in which complex models can be analyzed and compared with one another. Useful PDFs can also be calculated in spaces defined by small numbers of singular vectors or bred vectors (Miller and Ehret, 2002, Toth and Kalnay, 1993 and references therein). We have found in the course of our earlier work that taking advantage of this low dimensional structure to develop insight into the evolution of PDFs of stochastic models depends critically on our understanding of the qualitative dynamics of the system, in this case the characterization of attracting sets and their stability.

Optimized methods require accurate knowledge of the statistics of the errors in the model and the data. It is therefore an objective to understand in detail the sensitivity of the data assimilation scheme to the details of the defining error estimates.

APPROACH

The basic assumptions underlying data assimilation methods in use or proposed are known to be false to some degree. We plan to study the consequences of these assumptions by constructing a hierarchy of schemes with decreasing reliance on ad hoc assumptions. It is our guiding philosophy that the best way to learn how to design and implement the most economical methods that meet our needs is to begin by implementing methods which are as close to optimal as possible. From that point, we can quantify the loss of accuracy and the saving of resources associated with each simplification of the model or the data assimilation scheme.

Work is proceeding toward a theoretical basis for the next generation of data assimilation methods in which randomness and nonlinearity must be taken into account. To this end, we are applying tools from stochastic differential equations and from dynamical systems theory. Since our model systems are characterized by high dimensional state spaces, Monte Carlo methods must be used to study the behavior of the stochastic systems.

In systems that are well described by linear models, one may safely assume the PDF of the stochastic system to be Gaussian. This also holds for weakly nonlinear systems, but not for strongly nonlinear ones. Nearly all data assimilation systems implemented and proposed are based on assumptions of near-linearity. In the course of our work we have implemented linearized methods for highly nonlinear systems and found situations in which they succeed, and others in which they fail; some examples can be found in Miller et al. (1999). In recent work, we have found that linearized methods can be successful in highly nonlinear regimes, such as eddy formation in the Kuroshio off the coast of Japan (Vernieres, 2006). The same system, however, was not successful in predicting the decay of the eddy. Systems such as this one can provide essential guidance in the study of data assimilation in strongly nonlinear systems.

The theory of nonlinear filtering provides a framework in which problems of data assimilation with nonlinear models and non-Gaussian noise sources can be treated (see, e.g., Miller et al., 1999). In the case of linear models and Gaussian noise sources, this theory reduces to the familiar Kalman filter. In the formal theory of nonlinear filtering, the final result is not a single model state vector or trajectory in state space, but a PDF defined as a scalar function of the state variables and time. From this PDF, the mean, median, mode, or other statistic can be computed for use as the working estimate of the state

of the system, along with the desired confidence intervals. The assignment of confidence limits corresponds in the case of a group of particles in physical space to drawing contours in the spatial domain which can be expected to define a region which contains, say, 90% of the particles.

We are now in the process of investigating the qualitative behavior of more complex and relevant systems than the schematic ones with which we worked in the past (Miller et al., 1999, Miller and Ehret, 2002). At this moment, our work is focused on models of the north Pacific, with focus on the Kuroshio. The simplest of these is a regional quasigeostrophic (QG) model that reproduces the observed bimodality. It operates on a state space with several thousand dimensions. This is two orders of magnitude greater than that of earlier schematic models, and, for this reason alone, presents significant technical challenges.

We now have a basis of comparison with more complex models, up to and including eddy resolving primitive equation models of the north Pacific. We are now in the process of applying our methods from dynamical systems and stochastic calculus to a suite of models, in order to understand propagation of errors and the evolution of the PDF arising from random initial and boundary conditions in a state space of workable dimension.

Many different models, based on fundamentally different physical assumptions, exhibit the observed bimodality of the Kuroshio in some form. We are now in the process of performing comparisons among regional and basin scale models with differing physical and computational features, as well as comparisons of all of these models to observed data in order to determine a basis for distinction between the physical mechanisms in the different models.

Professor James G. Richman of the College of Oceanic and Atmospheric Sciences at Oregon State is working with us on the dynamical analysis of the Kuroshio. He is also working with us on error estimates (Richman et al., 2005), analysis of time series, and on comparison of ocean general circulation model results to data and to other models. Technical support for this project is provided by Ms. Laura Ehret.

WORK COMPLETED

We have developed and tested a suite of models from regional to basin scale suitable for rigorous mathematical treatment, assimilation of real data and comparison to detailed general circulation models, along with the ability to provide quantitative diagnostics.

We have completed our suite of models with the addition of a basin-scale model with eddy-resolving horizontal resolution and coarse vertical resolution similar to that described by Qiu and Miao (2000). Our version of this model is based on the baroclinic QG regional model implemented by Vernieres (2006).

RESULTS

The horizontal divergence of the velocity field of the $1/10^o$ primitive equation (PE) model of the Kuroshio off the coast of Japan (kindly provided by Dr. J. McClean; model as described in Smith et al., 2000) is typically of order 0.1, appropriately scaled, the same level as the Rossby number. Past

experience with QG models has shown that this level to be roughly the limit at which QG models give useful quantitative results.

Preliminary results from extension of our regional model to basin scale (see figure 1) seem to show that the regional and basin scale models are internally consistent.
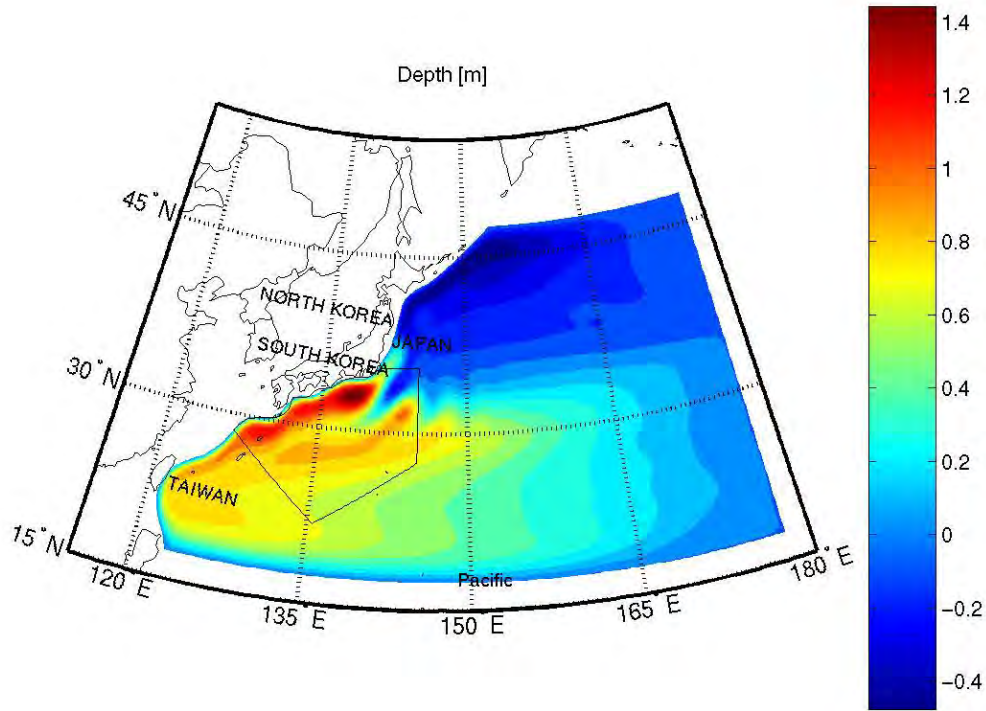

Figure 1. Result of a long integration of a two-layer QG model on a curvilinear grid, driven by the Hellerman and Rosenstein (1983) winds. Colors show sea surface height anomaly in meters. Inset region south of Japan is the domain of the limited area model.

IMPACT/APPLICATIONS

Major weather centers, including the US National Center for Environmental Prediction (NCEP) and the European Center for Medium-Range Weather Forecasting (ECMWF) now use ensemble methods for operational forecast validation; see Molteni et al. (1996), Toth and Kalnay (1993). Our work on Monte-Carlo methods should provide enhanced capability for validation of forecasts of the ocean and atmosphere, in addition to application to data assimilation. Our work on breeding modes and planned work on other schemes for ensemble generation should provide significant guidance in optimizing methods for generation of ensembles. Our work on dynamical analysis of models of the Kuroshio should lead to practical methods for identification of low-dimensional spaces in which efficient ensemble methods could be implemented.

We expect our the results of our inverse model of the Kuroshio to shed light on the importance of nonlinearity in ocean models; further, we expect that our work with comparisons among models and data for the Kuroshio will lead to greater insight into the intrinsic variability of basin-scale ocean circulation.

RELATED PROJECTS

"Assimilation of Coastal Radar Surface Current Measurements in Shelf Circulation Models." Work is in progress on the investigation of data assimilation systems for use with surface velocity data from coastal radar. This project is in collaboration with Professors Alexandre Kurapov, John Allen and Gary Egbert.

"Estimating the representation error of satellite and in situ data for data assimilation into ocean climate models." This project, a collaboration of the PI with Professor James Richman, is sponsored by the Joint Center for Satellite Data Assimilation.

REFERENCES

Hellerman, S. and Rosenstein, M., 1983: Normal monthly wind stress over the world ocean with error estimates. J. Phys. Oceanogr., 13, 1093-1104.

Miller, R.N. and L. Ehret (2002) Ensemble Generation for Models of Multi-modal Systems. *Monthly Weather Review*, **130**, 2313-2333.

Miller, R. N., E. F. Carter and S. T. Blue, 1999: Data assimilation into nonlinear stochastic models. *Tellus*, **51A**, 167-194**.**

Molteni, F., R. Buizza, T. N. Palmer and T. Petroliagis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Q. J. R. Meteorol. Soc.*, **122**, 73-119.

Richman, J. G., R. N. Miller and Y. H. Spitz, 2005: Error estimates for assimilation of satellite sea surface temperature data in ocean climate models. *Geophys. Res. Lett.* **32**, L18608, doi:10.1029/2005GL023591

Smith, R. D., Maltrud, M. E., Bryan, F. O. and Hecht, M. W., 2000: Numerical simulation of the north Atlantic at 1/10°. *J. Phys. Oceanogr.* **30,** 1532-1560.

Toth, Z. and E. Kalnay, 1993: Ensemble forecasting at NMC: the generation of perturbations. *Bull. Am. Meteorol. Soc.*, **74**, 2317-2330.

Vernieres, G., 2006: *Modeling Studies of Ocean Circulation Using Inverse Methods and Bifurcation Theory.* Doctoral dissertation, College of Oceanic and Atmospheric Sciences, Oregon State University. 148pp.

PUBLICATIONS

Miller, R. N. (2006) Topics in Data Assimilation : Stochastic Processes. *Physica D*, **230**, 17-26 doi:10.1016/j.physd.2006.07.015 [published, refereed]

Miller, R. N. (2007) Discussion: Subjective likelihood for the assessment of trends in the ocean's mixed layer depth. *Journal of the American Statistical Association*. [In press, refereed].

Miller, R.N. (2007) *Numerical Modeling of Ocean Circulation*. Cambridge University Press 252pp. . [book, published]